(2)

AD-A219 074

DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 90-0311

6a. NAME OF PERFORMING ORGANIZATION
University of Texas at Austin

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS (City, State, and ZIP Code)
Dept. of Electrical and Computer Engineering
Austin, Texas 78712-1084

7b. ADDRESS (City, State, and ZIP Code)
Bldg. 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL (If applicable)
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-86-0029

8c. ADDRESS (City, State, and ZIP Code)
Bldg. 410
Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO
61102F	2304	A1	

11. TITLE (Include Security Classification)
Comments on the Sensitivity of the Optimal Cost and the Optimal Policy for a Discrete Markov Decision Process (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
Sernik, Enrique L., and Marcus, S.I.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Reprint	FROM _____ TO _____	1989 September	10

16. SUPPLEMENTARY NOTATION
PROCEEDINGS OF THE 1989 ALLERTON CONFERENCE, September 27-29, 1989.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Sensitivity analysis, Markov decision processes, dynamic programming.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
The problem of characterizing the effects that uncertainties and/or small changes in the parameters of a model can have on optimal policies is considered. It is shown that changes in the optimal policy are very difficult to detect even for relatively simple models. By showing for a machine replacement problem modeled by a partially observed, finite state Markov decision process, that the infinite horizon, optimal discounted cost function is piecewise linear, we find formulas to compute the optimal cost and the optimal policy, thus providing a means for carrying out sensitivity analyses. Examples are presented to show the usefulness of the results.

Keywords: Algorithms, Stochastic control, dynamic programming, (KR)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. Marc Jacobs

22b. TELEPHONE (Include Area Code)
(202) 767-5027

22c. OFFICE SYMBOL
NM

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

COMMENTS ON THE SENSITIVITY OF THE OPTIMAL COST AND THE OPTIMAL POLICY FOR A DISCRETE MARKOV DECISION PROCESS*

ENRIQUE L. SERNIK and STEVEN I. MARCUS**
**Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

The problem of characterizing the effects that uncertainties and/or small changes in the parameters of a model can have on optimal policies is considered. It is shown that changes in the optimal policy are very difficult to detect even for relatively simple models. By showing for a machine replacement problem modeled by a partially observed, finite state Markov decision process, that the infinite horizon, optimal discounted cost function is piecewise linear, we find formulas to compute the optimal cost and the optimal policy, thus providing a means for carrying out sensitivity analyses. Examples are presented to show the usefulness of the results.

Key words: Sensitivity analysis, Markov Decision Processes, Dynamic Programming.

1. INTRODUCTION

The problem of finding explicit descriptions and/or structural properties of optimal control laws and costs for partially observed (PO) stochastic control problems has received considerable attention in recent years (e.g., Ref. 1, 2). This is due in part to the computational advantages that have resulted from such descriptions (e.g., Ref. 3, 4, 5), and in part to the increasing interest in the design of adaptive control techniques aimed at overcoming changes in the optimal law caused by uncertainties and/or small changes in the parameters of the physical system being modeled (e.g., Ref. 6, 7, 8).

In this note we are intersted in finding how uncertainties and/or small changes in the parameters of the model affect the optimal policy and cost of a discrete PO Markov decision process (MDP). A simple example of such a process is that associated with a machine that produces items, namely a process that can be in either a "good" or in a "bad" state (corresponding to whether the machine produces or needs to be replaced). Since the state of the machine is monitored incompletely, this problem is converted to an equivalent completely observable (CO) MD problem (see e.g. Ref. 3, or Ref. 9, chapter 3), in which the conditional probability vector $\pi(t) = (\pi_1(t), \cdots, \pi_N(t))$, with $\pi_i(t)$ the probability that the machine is in state i ($i \in \{1, \cdots, N\}$) given past observations and actions, provides all the relevant information to select the control action at time t.

Even for this simple model, the effect that uncertainties on the parameters have on the optimal policy and cost is not easy to determine. The complication arises due to several reasons, including the following: (i) The optimal control action has to be specified for each of the (uncountably infinite number of) values of $\pi(t)$. This should be contrasted with the case of perfect observations where one need only compute the control for each of the finite number of values of the states; (ii) The control process for this production problem takes values in a finite set (namely one can let the machine produce, and either inspect or replace the machine), and so derivatives with respect to the control are not defined; (iii) It is well known (Ref. 1-9) that the optimal cost for the problem considered here satisfies a functional equation, which can be solved by using the Dynamic Programming (DP) algorithm (see Ref. 9, chapter 5). Computationally however, this is

* Research supported in part by the Air Force Office of Scientific Research under Grant AFOSR-86-0029, in part by the National Science Foundation under Grant ECS-8617860, and in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-86-C-0045.

Presented at the 1989 Allerton Conference

not an easy problem. The space in which $\pi(t)$ takes its values is discretized by means of a grid which has to be changed several times until one is certain that the result of (applying) the algorithm is actually a solution of the functional equation, and is not just an artificial result of the numerical discretization (see the examples below).

We investigate the dependence of the optimal cost and the optimal policy on any of the parameters of the model with two states before considering a more general model. The equations involved are scalar equations since $\pi(t)$ can be written as $\pi(t) = (1 - p(t), p(t))$, where $p(t)$ is the probability that the system is in the bad state at time t given past observations and actions; $p(t)$ will usually be denoted by p, omitting explicit dependence on t. The scalar formulation facilitates the determination of structural properties of the cost and the policy.

2. MODEL, NOTATION AND REVIEW OF PREVIOUS WORK

We denote by $\{x_t, t = 0, 1, \cdots\}$ the finite state MP associated with a machine that produces items; the state space is $X = \{0, 1\}$, also referred to henceforth as {good, bad}, respectively. Denote by $\{u_t, t = 0, 1, \cdots\}$ the control process; $u_t \in U = \{0, 1, 2\}$, also referred to as {produce, inspect, replace}, respectively. We associate a cost with each control action as follows: the cost associated with replacing the machine is denoted by R, and the cost associated with inspection by I. The cost associated with production is 0 if the machine is in the good state, and C if the machine is in the bad state. We assume that $0 < C < I < R$.

At time t one must decide whether to inspect the item produced or not, and whether to replace the machine or not. If the machine is replaced it will be in the good state at the end of period t. It is further assumed that no item was produced during that period. Inspection might be carried out to determine the state of the machine, but it will not imply a decision about whether to replace the machine. The observation process $\{y_t, t = 0, 1, \cdots\}$ takes values in $Y = \{0, 1\}$.

Assume that the probability vector $p^0 = (p_0^0, p_1^0)$ is given, where $p_i^0 = Pr\{x_0 = i\}$, $i = 0, 1$, and $\pi(0) = p^0$. The machine evolves according to transition probabilities $p_{ij}(u_t)$ defined by $p_{ij}(v) = Pr\{x_{t+1} = j / x_t = i, u_t = v\}$. Let $P(u_t)$, $u_t \in U$, be the transition probability matrices with entries $p_{ij}(u_t)$. The transition matrices $P(u_t)$, $u_t \in U$, are given by:

$$P(0) = P(1) = \begin{bmatrix} 1-\theta & \theta \\ 0 & 1 \end{bmatrix}, \quad P(2) = \begin{bmatrix} 1-\theta & \theta \\ 1-\theta & \theta \end{bmatrix}, \quad t = 0, 1, \cdots \tag{1}$$

where θ is the probability of machine failure in one time step.

The observation process is related to the state and the control processes by means of the conditional probabilities $q_{ik}(v) = P\{y_{t+1} = k / x_t = i, u_t = v\}$, with $q_{ik}(u_t)$ the entries of the observation matrices $Q(u_t)$, $u_t \in U$, given by:

$$Q(0) = Q(1) = Q(2) = \begin{bmatrix} q & 1-q \\ 1-q & q \end{bmatrix}, \quad t = 0, 1, \cdots \tag{2}$$

where $q \in [0.5, 1.0]$ is the probability of making a correct observation. The model is the one described by Ross in Ref. 10.

We are interested in the infinite horizon case, and the objective is to find an optimal admissible control policy that minimizes the expected discounted cost $J_g(p^0)$, given by:

$$J_g(p^0) = E_{p^0}\left[\sum_{t=0}^{\infty} \beta^t c(x_t, u_t)\right] \tag{3}$$

where $E_{p^0}[.]$ denotes conditional expectation with respect to p^0; β is the discount factor with $0 \leq \beta < 1$; $c(x_t, u_t)$ is the cost accrued when the machine is in state x_t and action u_t is selected; and $g = \{g_t\}_{t=1}^{\infty}$ is an admissible policy, that is, $\{g_t\}_{t=1}^{\infty}$ is a sequence of

Borel measurable maps $g_t:[0,1] \to U$ such that $u_t = g_t(p(t))$, $u_t \in U$, for $t = 0, 1, \cdots$. If no observations are available, u_t can still be written as $u_t = g_t(p(t))$, $u_t \in U$, and $g_t:[0,1] \to U$ for $t = 0, 1, \cdots$, where now $p(t)$ is the (aposteriori) probability that the machine is in state 1. This is because the expected cost can be expressed explicitly in terms of $p(t)$. However, in this case $\{g_t\}_{t=1}^{\infty}$ is a deterministic sequence since $p(t)$ depends only on p^0, which is given, and is updated from time t to time $t+1$ using the transition probabilities $P(.)$, also given. If $g_t(.) = g(.)$ for all values of t, the policy is said to be stationary (when computing optimal policies in the infinite horizon case, we need only consider stationary policies; see Ref. 9, p. 225).

Define $V_\beta(p) = \inf_g J_g(p)$. Then $V_\beta(p)$ is the expected cost accrued when an optimal policy is selected, given that the machine starts in the bad state with probability p, and future costs are discounted at rate β. It is well known (e.g. Ref. 9, 10) that $V_\beta(p)$ is the unique solution of:

$$V_\beta(p) = \min \{ C\,p + \beta \sum_{k=0}^{1} D(k,p,0)\, V_\beta(T(k,p,0)),$$

$$I + \beta \sum_{k=0}^{1} D(k,p,1)\, V_\beta(T(k,p,1)),$$

$$R + \beta \sum_{k=0}^{1} D(k,p,2)\, V_\beta(T(k,p,2)) \} \qquad (4)$$

where $T(k,p,v)$ is the updated probability that the system is in state 1, given that k was observed and the control applied was v. $T(k,p,v)$ is given by $T(k,p,v) = N(k,p,v)/D(k,p,v)$, with $N(k,p,v) = (N_1(k,p,v), N_2(k,p,v))$, $D(k,p,v) = \sum_j N_j(k,p,v)$, and where $N_j(k,p,v)$ represents the probability that the next state is j given that the outcome is k and the control applied is v (see Ref. 9, 10 for details).

We review some previous work associated with the following two special cases of the strictly PO (i.e., partial observations during production and during inspection) problem: **Case A**: Only two actions are considered, namely $U = \{0, 2\}$, and the state of the system is not observed (i.e., the state is 'completely unobserved') during production, that is, all the entries in $Q(0)$ equal 0.5; **Case B**: Now $U = \{0, 1, 2\}$, and the state of the system is not observed during production, but perfect observations (i.e., the state is 'completely observed') are obtained during inspection, so that $Q(0)$ is as in **Case A** and $Q(1)$ is the 2-dimensional identity matrix. We recall that these two cases are of interest since the completely unobserved (CU) and the completely observed (CO) cases repectively provide upper and lower bounds for the optimal value of the cost in the strictly PO case (Ref. 11).

The first structural results associated with the optimal policy and cost for the models described above were given by Ross (Ref. 10). Among several results, Ross gave necessary and sufficient conditions for the stationary policy "produce for all values of p" to be optimal. Ross also showed that (i) every optimal policy produces for all $p \in [0, \theta]$; and that (ii) it is optimal to replace for values of p near 1. Other results by Ross included sufficient conditions to verify the existence of optimal policies. The conditions were stated in terms of the optimal cost and were thus hard to verify. The characterization of the stationary optimal policies was done by White (Ref. 3), who showed that among the stationary optimal policies there is a smaller class of optimal policies, called structured policies, such that one need only look for structured policies when solving equation (4). Wang's work (Ref. 12) was aimed at showing that a structured policy (called 'control-limit policy' when only two actions are considered) is optimal for the two action, CU case. Wang also gave analytic expressions for computing the optimal cost and the optimal policy for this problem. Although these results can be used to show that the stationary optimal cost is piecewise linear, Wang did not do so, and unfortunately his results have been referred to primarily as a "computational procedure" (see e.g., Ref. 13). Wang stu-

died a more general model for the two action, CU case than the one being treated here, but he did not considered the case of three actions (closed loop). In a later work (Ref. 14), Wang generalized his results to the two action, CU N – dimensional ($N > 2$) case.

Let us point out that the previous results characterize the optimal policy and give some properties of the optimal cost function for the problems described above, but they do not give insight on what happens to the optimal policy or to the optimal cost if there is uncertainty in the knowledge of the parameters of the model, or if these parameters undergo (unexpected) small changes. Solving the problem again via Dynamic Programming may not be practical for the infinite horizon case in terms of computational effort, as will be illustrated in the examples below.

3. PIECEWISE LINEAR OPTIMAL COST

Motivated by our interest in determining the sensitivity of the optimal policy with respect to the parameters of the model, and since the study of the functional equation (4) provides little insight on how the optimal policy or the optimal cost change when the parameters of the model are subject to small changes, we focused our attention on the study of the DP (or successive approximations, Ref. 9, 10) algorithm used to solve equation (4). Specifically, we analyzed the algorithm given by:

$$V_\beta^1(p) = \min\{C\,p\ ,\ I\ ,\ R\ \}$$

$$V_\beta^n(p) = \min\{C\,p + \beta \sum_{k=0}^{1} D(k,p,0)\, V_\beta^{n-1}(T(k,p,0)),$$

$$I + \beta \sum_{k=0}^{1} D(k,p,1)\, V_\beta^{n-1}(T(k,p,1)),$$

$$R + \beta \sum_{k=0}^{1} D(k,p,2)\, V_\beta^{n-1}(T(k,p,2))\ \} \tag{5}$$

where n represents the iteration, and $V_\beta^n(p)$ is the minimal cost that can be obtained starting in state 1 with probability p and proceeding for n stages with costs discounted by a factor β. Because of the relative simplicity of algorithm (5), and since from the theory of contraction mappings it is guaranteed that algorithm (5) converges uniformly to the unique solution $V_\beta(p)$ of equation (4) as $n \to \infty$ (see e.g. Ref. 2), we were able to prove the piecewise linearity of the optimal cost function, and obtained analytic expressions to compute the optimal cost and the optimal policy. We show this next.

Consider first Case A described above. When $q = 0.5$ we have $D(k,p,0) = D(k,p,2) = \frac{1}{2}$, $T(k,p,2) = \theta$, and $T(k,p,0)$ becomes (Ref. 9):

$$T(k,p,0) = p(1-\theta) + \theta \equiv Tp \tag{6}$$

Thus, Tp satisfies $Tp > p$ for $p \in [0,1)$, with unique fixed point $p = 1$. From algorithm (5), denote by $\bar{P}_k(p)$ the function generated by $\bar{P}_k(p) \equiv C\,p + \beta \bar{P}_{k-1}(Tp)$, $k = 2, 3, \cdots$, with $\bar{P}_1(p) \equiv C\,p$, and denote by $\bar{R}_k$ the function generated using $\bar{R}_k \equiv R + \beta \bar{P}_{k-1}(\theta)$, $k = 2, 3, \cdots$, with $\bar{R}_1 \equiv R$. By applying recursively (5) one obtains that:

$$\bar{P}_k(p) = C\,p \sum_{i=0}^{k-1} \beta^i (1-\theta)^i + C \sum_{i=1}^{k-1} \beta^i (1-(1-\theta)^i) \tag{7}$$

$$\bar{R}_k = R + C \sum_{i=1}^{k-1} \beta^i (1-(1-\theta)^i) \tag{8}$$

Since $C < R$ by hypothesis, and $p < 1$, the first iteration of (5) gives the policy "produce for all $p \in [0,1]$". As k increases, if $\bar{P}_k(p) \neq \bar{R}_k$ for all $k \in N$, and all $p \in [0,1]$, one obtains Ross' result mentioned above, namely, that "produce for all $p \in [0,1]$" is the

stationary, infinite horizon optimal policy. If on the contrary $\bar{P}_k(p) = \bar{R}_k$ for some $k \in N$ and for some $p < 1$, call it α_k, then the optimal cost at iteration k will be specified by:

$$V_\beta^k(p) = \begin{cases} C p \sum_{i=0}^{k-1} \beta^i (1-\theta)^i + C \sum_{i=1}^{k-1} \beta^i (1-(1-\theta)^i) & p \in [0, \alpha_k) \\ R + C \sum_{i=1}^{k-1} \beta^i (1-(1-\theta)^i) & p \in [\alpha_k, 1] \end{cases} \tag{9}$$

with optimal policy "produce for $p \in [0, \alpha_k)$ and replace for $p \in [\alpha_k, 1]$". The point we want to make is the following: for $n = k+1$ (and similarily for subsequent iterations) in order to compute $V_\beta^{k+1}(p)$ we need to perform a minimization in an interval of the form $[0, b)$, $0 < b < 1$, requiring the evaluation of $V_\beta^k(Tp)$, $p \in [0, b)$. But since $Tp > p$ for $p < 1$, the result of the minimization only specifies the cost function in an interval of the form $[0, T^{-1}b)$ at iteration $k+1$. This implies the following:

(i) Since $p = 1$ is the unique fixed point of $T^{-1}p = (p-\theta)/(1-\theta)$, and $T^{-1}p$ continues to decrease the size of the interval of the form $[0, b)$ as $k \to \infty$, there is an iteration, call it l, for which this interval is smaller than $[0, \theta)$ (say $[0, \gamma), \gamma < \theta$). Since $T^{-1}p < 0$ for $p \in [0, \gamma)$, the cost specified in $[0, \gamma)$ in iteration l will not enter into the computation of $V_\beta^{l+1}(p)$, $V_\beta^{l+2}(p)$, $\cdots$.

(ii) To specify the cost function on the remainder of the interval, namely $[T^{-1}\alpha_k, 1]$, algorithm (5) requires a minimization of the form:

$$\min \{C p + \beta \bar{R}_k, \bar{R}_{k+1}\} \tag{10}$$

Note that in (10) we are comparing a function associated with the produce action, given by an affine function of p (referred to henceforth simply as a 'line segment' in order to facilitate the exposition), and $\bar{R}_{k+1}$, a constant. If $C p + \beta \bar{R}_k$ is smaller than $\bar{R}_{k+1}$ in (10) (for some value of $p \in [T^{-1}\alpha_k, 1]$), a new line segment will appear in the description of the cost function (for the interval $p \in [T^{-1}\alpha_k, \alpha_{k+1})$). The point here is that with the exception of the line segment shown in (9), all the line segments appearing in the description of the cost function come from a minimization of the form (10).

In other words, from (ii), at each iteration of (5), and independently of the number of line segments already describing the cost function in that iteration, there is at most one new line segment appearing in the description of the optimal cost function, and from (i), for k large enough the line segments are also leaving the problem (meaning they no longer appear in the cost function) under the action of $T^{-1}p$.

(iii) In addition, observe that the line segment that specifies the cost function at iteration (say) k in the interval $[a, b)$, $0 < a < b < 1$, specifies (with formula updated by the iterative procedure) the cost function at iteration $k+1$ in the interval $[T^{-1}a, T^{-1}b)$, and since for $0 < a < b < 1$ we have that $T^{-1}b - T^{-1}a = (b-a)/(1-\theta) > b-a$, the line segments leaving the problem have finite nonzero length.

Finally, if we call α^* the limit as $k \to \infty$ of α_k (whenever it exists), then there is a finite natural number m_{α^*} such that $T^{-m_{\alpha^*}}\alpha^*$ is less than zero (this is clear because $T^{-1}p < p$ for $0 \le p < 1$, and the fact that $p = 1$ is the only fixed point of $T^{-1}p$). The same is true for each of the α_k, $k \in N$. This observation, together with remarks (i), (ii) and (iii) above, means that all the line segments that appear in the description of the cost function disappear in a finite number of iterations.

We introduce the following notation. Let $W_\beta^k(p) \equiv V_\beta^k(p)|_{[0,\alpha_k)}$, that is, $W_\beta^k(p)$ denotes the restriction of the optimal cost function at iteration k to the interval $[0, \alpha_k)$. Observe that in this notation $\bar{R}_k$ can be interpreted as the restriction of $V_\beta^k(p)$ to the interval $[\alpha_k, 1]$. Let $W_\beta(p)$ and $\bar{R}$ be the limits (whenever they exist) of $W_\beta^k(p)$ and $\bar{R}_k$ respectively, as $k \to \infty$. Then $W_\beta(p)$ (respectively $\bar{R}$) denotes the restriction of the infinite horizon, optimal discounted cost function $V_\beta(p)$ to the interval $[0, \alpha^*)$ (respec-

tively $[\alpha^*, 1]$).

One of two things can happen so that convergence is achieved: either (a) The line segments enter into the problem at a higher rate than that at which they disappear from the problem, and hence the line segments accumulate, meaning that in the limit as $k \to \infty$, $W_\beta(p)$ will not be described by a finite number of line segments; or else (b) The rate at which the line segments enter the problem is the same as that at which they leave the problem, and hence a finite number of line segments completely describe $W_\beta(p)$. We have the following Proposition.

Proposition 3.1: For the open loop model described above in *Case A* (two actions, i.e., U = {produce, replace}, and the state of the system is CU during production), $W_\beta(p)$, associated with the stationary, infinite horizon optimal policy is piecewise linear. Since $\bar{R}$ is constant, this means that the infinite horizon, optimal cost function $V_\beta(p)$ is piecewise linear.

Proof: By Lemma 3.2 in Ref. 10, we have that every optimal policy produces for all $p \in [0, \theta]$. Hence we consider the following two cases:

(i) If $\alpha^* = \theta$, then $T^{-1}\alpha^* = 0$, and the claim here is that the optimal cost $V_\beta(p)$ associated with the stationary, infinite horizon optimal policy "produce for $p \in [0, \theta]$ and replace for $p \in (\theta, 1]$", is given by:

$$V_\beta(p) = \begin{cases} Cp + \beta\bar{R} & p \in [0, \theta] \\ \bar{R} & p \in (\theta, 1] \end{cases} \tag{11}$$

For if not, assume that $W_\beta(p)$ is an arbitrary limit of piecewise linear functions. Denote by $f(p)$ the optimal cost function that is an arbitrary limit of piecewise linear functions for $p \in [0, \alpha^*]$, and a constant for $p \in (\alpha^*, 1]$. Then from (4) we have that: $f(p) = \min\{Cp + \beta f(Tp), R + \beta f(\theta)\}$. That is, by Lemma 3.2 in Ref. 10, and for $0 \le p \le \alpha^* = \theta$, we have $f(p) = Cp + \beta f(Tp)$. Since $f(Tp) = R + \beta f(\theta)$ is constant (say K) for $0 \le p \le \alpha^* = \theta$, we have $f(p) = Cp + K$ for $p \in [0, \alpha^*]$. Therefore, if $\alpha^* = \theta$, the infinite horizon, optimal cost is given by (11), which means that there is only one line segment describing $W_\beta(p)$.

(ii) Since the case for which $\alpha^* = 1$ was considered by Ross in Ref. 10, assume that $\theta < \alpha^* < 1$, and that convergence takes place as described in (a) above. As explained in remark (ii) above, the line segment that appeared in the problem most recently specifies the cost function in the interval $[T^{-1}\alpha^*, \alpha^*)$. Since $\alpha^* > T^{-1}\alpha^*$, this line segment has finite nonzero length. Now assume that for $p \in [0, T^{-1}\alpha^*)$ the cost function is arbitrary. Denote by $f(p)$ the optimal cost function that is an arbitrary limit of piecewise linear functions for $p \in [0, T^{-1}\alpha^*)$, an affine function of p for $p \in [T^{-1}\alpha^*, \alpha^*)$, and constant for $p \in [\alpha^*, 1]$. From (4), $f(p)$ is given by $f(p) = \min\{Cp + \beta f(Tp), R + \beta f(\theta)\}$. Since for $p \in [T^{-1}(T^{-1}\alpha^*), T^{-1}\alpha^*)$ the optimal policy is to produce, $f(p) = Cp + \beta f(Tp)$ for $p \in [T^{-1}(T^{-1}\alpha^*), T^{-1}\alpha^*)$. But for $p \in [T^{-1}(T^{-1}\alpha^*), T^{-1}\alpha^*)$ we have $Tp \in [T^{-1}\alpha^*, \alpha^*)$, and so $f(Tp)$ is the above mentioned segment in $[T^{-1}\alpha^*, \alpha^*)$ with finite (nonzero) length, which in turn means that $f(p) = Cp + \beta f(Tp)$ is also an affine function of p, and since $T^{-1}\alpha^* > T^{-1}(T^{-1}\alpha^*)$, it also has a finite (nonzero) length. By remark (iii) above, we have that $\alpha^* - T^{-1}\alpha^* > T^{-1}\alpha^* - T^{-1}(T^{-1}\alpha^*)$. Therefore, continuing the procedure just described, we can see that there is a uniform lower bound (different from zero) for the length of all the line segments which is independent of the iteration because $T^{-1}p$ is independent of the iteration. Taking into account the previous observations, we conclude that this uniform lower bound for the length of the line segments implies an upper bound in the number of line segments describing $V_\beta(p)$. **QED**

At this point the following remarks are in order:

(i) To the best of our knowledge, the piecewise linearity of the optimal cost function in the infinite horizon model for the cases described above has not been reported previously (see for example all the references cited so far, and recent reviews like Ref. 16). Its usefulness will be apparent in the sequel.

(ii) Although the two action, CU model may be of limited interest in practice (namely, it may only be used in some replacement models in which the equipment is subject to breakdowns but no measurements are available, and therefore only open loop control is applicable), its importance here resides in that the insight obtained by its study allowed us to perform a similar analysis (and also prove the piecewise linearity of the optimal cost function) for the three action, closed loop case.
(iii) The analysis of the successive approximations algorithm for the case of three actions is not trivial. As pointed out in Ref. 15, p. 29, a policy can appear at any time during the iterative procedure, yet fail to be optimal for the infinite horizon case. Furthermore, it might also happen that a policy appears during some iteration, does not appear in the (e.g., one hundred) subsequent iterations, and reappears later (or never reappears), meaning that a policy structure which is not any longer optimal after some finite iteration $k \in N$, cannot be eliminated as suboptimal, and so, estimation of the minimum number of iterations required (for example in algorithm (5)) to guarantee that an optimal policy for the n^{th} (finite) horizon is also the optimal policy for the infinite horizon case, remains an open problem (Ref. 15).

For the sake of brevity, we will not go into the details of the three action, closed loop problem, since it requires a lenghty analysis of the successive approximations algorithm (5). However, we have studied the algorithm (5) and shown that some policy structures cannot occur at all during the iterative procedure (e.g., there is not a "produce-inspect" policy), and that the occurrence of others do not affect the stationary, infinite horizon policy structure (Ref. 17).

With these results we were able to show that the infinite horizon, optimal cost function in the three action, closed loop problem is piecewise linear, and were able to develop analytic expressions (equivalent to those of Wang in Ref. 12 for the two action, CU case, and new ones for the three action, closed loop case) for the costs and the structured optimal policies for **Cases A** and **B** (namely, policies that as a function of p have the characterization: "there exist three numbers ρ_i, $i = 1,2,3$, $\theta \leq \rho_1 \leq \rho_2 \leq \rho_3 \leq 1$, such that it is optimal to produce for $0 \leq p < \rho_1$ ($0 \leq p \leq \rho_1$ if $\rho_1 = \theta$) and $\rho_2 \leq p < \rho_3$, it is optimal to inspect for $\rho_1 \leq p < \rho_2$, and it is optimal to replace for $\rho_3 \leq p \leq 1$"; see Ref. 3, 10 for detailed analysis). This in turn allows us to avoid the computational burden described in Section 2 when solving the control problem: in particular, the results allow us to perform a sensitivity analysis of the optimal policy with respect to any of the parameters of the problem, as will be illustrated in the examples ahead.

Once it is established that only a finite number of line segments is required to completely describe the infinite horizon optimal cost function, algorithm (5) can be used to find analytic expressions to compute the cost and the policy structure. Since the importance of these formulas reside in their use to perform sensitivity analyses of the optimal cost and policy, we illustrate this next with some examples.

4. EXAMPLES

Example 1: Consider the closed loop problem with the following data: $\beta = 0.985$, $\theta = 0.1$, $C = 4.0$, $I = 5.56$ and $R = 10.0$. The stationary, infinite horizon optimal policy is: "produce for $p \in [0, 0.5145193)$ and $p \in [0.5369462, 0.6070793)$, inspect for $p \in [0.5145193, 0.5369462)$ and replace for $p \in [0.6070793, 1]$", and the associated optimal cost is given by:

$$V_\beta(p) = \begin{cases} 23.33618p + 151.88782 & p \in [0.0000000, 0.0864824) \\ 21.81182p + 152.01965 & p \in [0.0864824, 0.1778342) \\ 20.09230p + 152.32544 & p \in [0.1778342, 0.2600507) \\ 18.15262p + 152.82985 & p \in [0.2600507, 0.3340457) \\ 15.96460p + 153.56075 & p \in [0.3340457, 0.4006411) \\ 13.49645p + 154.54959 & p \in [0.4006411, 0.4605770) \\ 10.71230p + 155.83191 & p \in [0.4605770, 0.5145193) \\ 7.57168p + 157.44782 & p \in [0.5145193, 0.5369462) \\ 7.54600p + 157.46161 & p \in [0.5369462, 0.5634215) \\ 4.00000p + 159.45950 & p \in [0.5634215, 0.6070793) \\ 161.88782 & p \in [0.6070793, 1.0000000] \end{cases} \quad (12)$$

Observe that equation (12) is a closed form formula for the infinite horizon discounted cost. Also, note that there are 7 line segments describing the optimal cost function for $p \in [0.0, 0.5145193)$, and 2 line segments describing the optimal cost function for $p \in [0.5369462, 0.6070793)$. Now change θ from 0.1 to 0.10005. The optimal policy structure now changes from 4 to 2 regions. In this case $\alpha^* = 0.60721$, and the optimal cost is given by:

$$V_\beta(p) = \begin{cases} 23.3210p + 151.9233 & p \in [0.00000, 0.08713) \\ 21.7959p + 152.0562 & p \in [0.08713, 0.17846) \\ 20.0754p + 152.3632 & p \in [0.17846, 0.26065) \\ 18.1346p + 152.8691 & p \in [0.26065, 0.33463) \\ 15.9452p + 153.6017 & p \in [0.33463, 0.40120) \\ 13.4753p + 154.5927 & p \in [0.40120, 0.46111) \\ 10.6890p + 155.8774 & p \in [0.46111, 0.51502) \\ 7.5458p + 157.4962 & p \in [0.51502, 0.56355) \\ 4.0000p + 159.4945 & p \in [0.56355, 0.60721) \\ 161.9233 & p \in [0.60721, 1.00000] \end{cases} \quad (13)$$

Note how a relatively small change in the value of θ resulted in a significant change in the optimal policy structure. To the best of our knowledge, changes in the optimal policy due to such small changes on the parameters of the model could not be studied before, because the necessity of discretization often does not permit high confidence in the results obtained by following the DP algorithm. We are able to study small changes in the parameters of the model because of the analytic expressions found as a consequence of the piecewise linearity of the optimal cost.

Now consider the case when θ changes to $\theta = 0.09$. The optimal policy is: produce for $p \in [0, 0.4874221)$ and $p \in [0.5772912, 0.5773791)$, inspect for $p \in [0.4874221, 0.5772912)$, replace for $p \in [0.5773791, 1]$, and the optimal cost is given by:

$$V_{\beta}(p) = \begin{cases} 25.71507p + 143.96777 & p \in [0.0000000, 0.0080949) \\ 24.22611p + 143.97982 & p \in [0.0080949, 0.0973664) \\ 22.56497p + 144.14156 & p \in [0.0973664, 0.1786034) \\ 20.71174p + 144.47256 & p \in [0.1786034, 0.2525291) \\ 18.64421p + 144.99467 & p \in [0.2525291, 0.3198015) \\ 16.33760p + 145.73232 & p \in [0.3198015, 0.3810193) \\ 13.76427p + 146.71281 & p \in [0.3810193, 0.4367276) \\ 10.89336p + 147.96661 & p \in [0.4367276, 0.4874221) \\ 7.69048p + 149.52777 & p \in [0.4874221, 0.5772912) \\ 4.00000p + 151.65825 & p \in [0.5772912, 0.5773791) \\ 153.96777 & p \in [0.5773791, 1.0000000] \end{cases} \quad (14)$$

The optimal policy for this example still has four regions if θ changes to any value in $[0.09, 0.10)$. However, note that the number of line segments describing the optimal cost function changes for different values of θ. Also, note the size of the interval for which the line segment with formula $4.0p + 151.65825$ is specified in (23). We compared these results with those given by the successive approximations algorithm. The results were very difficult to obtain by using the successive approximations algorithm. Unless one knows in advance the structure of the stationary, infinite horizon optimal policy, it is very difficult to decide when the optimal policy has been reached (and hence to decide when to stop the computational procedure), even after several choices of the grid have been tested (with the corresponding time consumption involved).

It is clear that the same kind of analysis carried out here can be done for any of the other parameters of the problem (i.e., β, R, I and C). Thus the equations found in this work can be used to obtain insight in the way the system responds to uncertainties, and therefore adaptive policies can be designed (for example, to modify the value of some of the parameters to compensate for an undesired change in some other parameters) so that the system continues to perform in a preselected satisfactory way.

Example 2: Now consider the open loop case. Let $\beta = 0.9999$, $\theta = 0.1$, $C = 4.0$ and $R = 10.0$. Using algorithm (5) with a grid of 1001 points in the interval $[0,1]$, one obtains: for $n = 1000$, $\alpha^* = 0.609$, $\bar{R} = 2272.10$; for $n = 10000$, $\alpha^* = 0.588$, $\bar{R} = 15077.17$; and for $n = 15000$, $\alpha^* = 0.601$, $\bar{R} = 18528.28$. We note that the last case mentioned above took 52 minutes of CPU (as compared to less than a second when using the expressions derived here, for the same computer and computer load), and although the values obtained may suffice when solving the (initial) control problem (actual values, obtained with the analytical expressions, are $\bar{R} = 23882.62$ and $\alpha^* = 0.597167$) it is apparent that a sensitivity analysis would be not only expensive in terms of computer time (this is so for any computational algorithm since p takes uncountably many values, and we are dealing with the infinite horizon problem), but also hard to perform in the sense of detecting the actual effect of the uncertainties on the optimal cost and policy.

5. CONCLUSIONS

The analysis of the successive approximations algorithm used to solve the functional equation satisfied by the optimal cost associated with the problems described in Section 3, allowed us to prove the piecewise linearity of the optimal cost function, and to develop analytical expressions to compute both the infinite horizon optimal cost and the stationary, infinite horizon optimal policy structure. These results, in turn, permit the perfor-

mance of a sensitivity analysis for the optimal cost and policy with respect to any of the parameters of the problem.

The examples in Section 4 suggest that for the study of changes in the optimal policy due to small changes in the parameters of the model, better results can be obtained if structural properties of the policies and the cost are taken into account in the design of computational procedures. Since the strictly PO case is difficult to treat analytically, the study of structural properties of the optimal cost and the optimal policy for special cases of the strictly PO case, like those considered here, is justified as a way to approach the strictly PO problem.

The development of similar results for more complex problems like the study of the strictly PO case, higher dimensional models, and the average cost case, is currently being investigated.

6. REFERENCES

[1] Hopp, W. J., *Sensitivity Analysis in Discrete Dynamic Programming*, Journal of Optimization Theory and Applications, Vol. 56, No. 2, February 1988, p. 257-269.

[2] Ohnishi, M., Kawai, H. and Mine, H., *An Optimal Inspection and Replacement Policy Under Incomplete State Information*, European Journal of Operatonal Research, Vol. 27, 1986, p. 117-128.

[3] White, C. C., *Optimal Inspection and Repair of a Production Process Subject to Deterioration*, Journal of the Operational Research Society, Vol. 29, No. 3, 1978, p. 235-243.

[4] White, C. C., and El-Deib, H. K., *Parameter Imprecision in Finite State, Finite Action Dynamic Programs*, Operations Research, Vol. 34, No. 1, Jan/Feb 1986, p. 120-129.

[5] White, C. C., Thomas, L. C., and Scherer, W. T., *Reward Revision for Discounted Markov Decision Problems*, Operations Research, Vol. 36, No. 6, Nov/Dec 1985, p. 1299-1315.

[6] Hernandez-Lerma, O., and Marcus, S. I., *Adaptive Policies for Discrete-Time Stochastic Control Systems with Unknown Disturbance Distroutions*, Systems & Control Letters, Vol. 9, 1987, p. 307-325.

[7] Hernandez-Lerma, O., and Marcus, S. I., *Adaptive Control of Discrete Discounted Markov Decision Chains*, Journal of Optimization Theory and Applications, Vol. 46, No. 2, June 1985, p. 227-235.

[8] Hernandez-Lerma, O., and Marcus, S. I., *Adaptive Control of Markov Processes with Incomplete State Information and Unknown Parameters*, Journal of Optimization Theory and Applications, Vol. 52, No. 2, February 1987, p. 227-241.

[9] Bertsekas, D. P., *Dynamic Programming*, Prentice Hall, Englewood Cliffs, New Jersey, 1987.

[10] Ross, S. M., *Quality Control Under Markovian Deterioration*, Management Science, Vo'. 17, No. 9, 1971, p. 587-596.

[11] Astrom, K. J., *Optimal Control of Markov Processes with Incomplete State Information*, Journal of Mathematical Analysis and Applications, Vol 10, 1965, p. 174-205.

[12] Wang, R. C., *Computing Optimal Control Policies - Two Actions*, Journal of Applied Probability, Vol. 13, 1976, p. 826-832.

[13] Monahan, G. E., *A Survey of Partially Observable Markov Decision Processes: Theory, Models and Algorithms*, Management Science, Vol. 28, No. 1, Jan. 1982, p. 1-16.

[14] Wang, R. C., *Optimal Replacement Policy with Unobservable States*, Journal of Applied Probability, Vol. 14, 1977, p. 340-348.

[15] Federgruen, A., and Schweitzer, P. J., *Discounted and Undiscounted Value-Iteration in Markov Decision Problems: A Survey*, in Dynamic Programming and its Applications. edited by M. Puterman, Academic Press, 1979, p. 23-52

[16] White, C. C., and White, D. J., *Markov Decision Processes*, European Journal of Operational Research, Vol. 39, 1989, p. 1-16.

[17] Sernik, E. L., and Marcus, S. I., *Sensitivity of the Optimal Cost and Policy for a Discrete Markov Decision Process*, submitted for publication.